BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37614
doi: 10.3897/biss.3.37614

OPEN ACCESS

Conference Abstract

From Field Observations and Plant Specimens to a Trans-continental Knowledge Base: Efficient, semantically rich integration of highly heterogeneous plant phenological data

Brian J. Stucky[‡], John Deck[§], Ramona L. Walls[|,¶], Robert Guralnick[‡]

‡ Florida Museum of Natural History, University of Florida, Gainesville, FL, United States of America
§ University of California at Berkeley, Berkeley, United States of America
| Bio5 Institute, University of Arizona, Tucson, AZ, United States of America
¶ CyVerse, Tucson, AZ, United States of America

Corresponding author: Brian J. Stucky (stucky.brian@gmail.com)

Received: 22 Jun 2019 | Published: 02 Jul 2019

Citation: Stucky B, Deck J, Walls R, Guralnick R (2019) From Field Observations and Plant Specimens to a Trans-continental Knowledge Base: Efficient, semantically rich integration of highly heterogeneous plant phenological data. Biodiversity Information Science and Standards 3: e37614. https://doi.org/10.3897/biss.3.37614

Abstract

Ideally, an information system that automates the integration of disparate datasets should be able to minimize the loss of information from any one dataset, achieve computational complexity suitable for working with large datasets, be flexible enough to easily incorporate new data sources, and produce output that is easily analyzed and understood by data users. Achieving all of these goals within highly heterogeneous and highly complex data domains is a major challenge. In this talk, we present the results of our recent efforts to develop such a system for data about plant phenology. Our data integration system, which is built around the Plant Phenology Ontology, currently supports semantically fine-grained integration of phenological data from both field observations and herbarium specimens. We show that even with a heavily axiomatized ontology and sophisticated, machine-reasoning-based data analysis, it is possible to implement a high-throughput data integration pipeline capable of processing millions of individual records in a matter of minutes while running on modest, server-class hardware. Success requires careful ontology design and judicious

© Stucky B et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

application of machine reasoning techniques. We also discuss some of the many challenges that remain for designing efficient, general-purpose data integration systems.

Keywords

ontology, data integration, machine reasoning, plant phenology

Presenting author

Brian J. Stucky

Presented at

Biodiversity_Next 2019